Statistical Inference - Simulation

Joy SN

January 3, 2019

Mean and Variance of Sample Mean for Exponential Distribution

Investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution is for 1000 simuulation of averages of 40 exponentials. Lamda is set to 0.2.

1. Packages required

```
library(ggplot2)
```

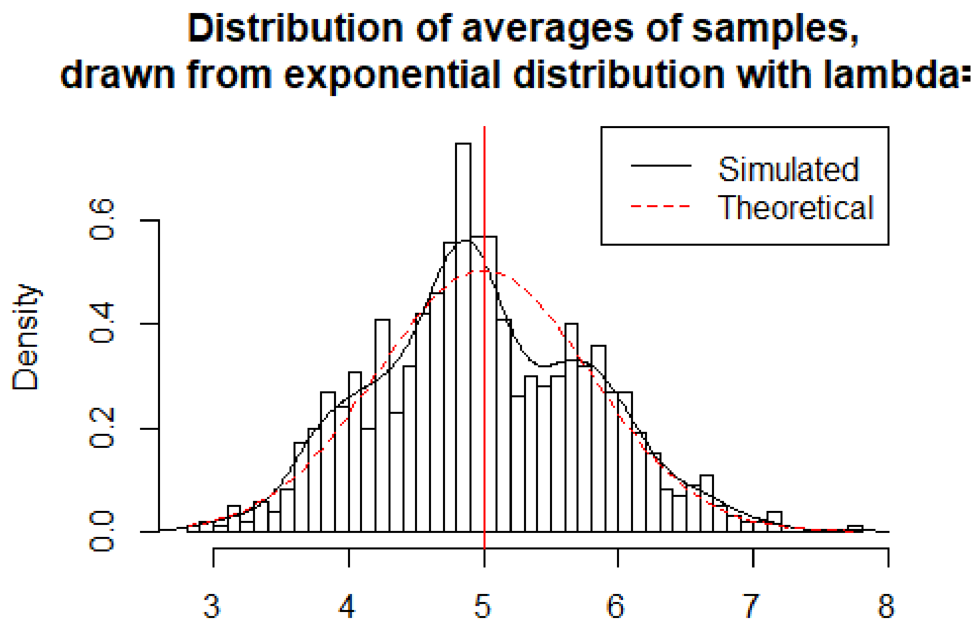
2. Data Setup

```
set.seed(19)
lambda <- 0.2
n <- 40 # size
simulations <- 1000 # number of simulations
simData <- replicate(simulations, rexp(n, lambda))
meanData <- apply(simData, 2, mean)
rowMeanData <- rowMeans(matrix(data = simData, nrow = simulations, ncol = n))
```

Where the distribution is centered at and compare it to the theoretical center of the distribution.

```
theoryMean <- 1/lambda
simulatedMean <- mean(rowMeanData) # Mean
rbind(simulatedMean, theoryMean)

##                   [,1]
## simulatedMean 4.991311
## theoryMean    5.000000

hist(rowMeanData, breaks=50, prob=TRUE, main="Distribution of averages of
samples,
     drawn from exponential distribution with lambda=0.2",
     xlab="")

# Plot the density curve for the means
lines(density(rowMeanData))

# Add 'theoretical center of distribution' for comparison
abline(v=1/lambda, col="red")
```

```r
# Add 'theoretical density for sample means' for comparison
xfit <- seq(min(rowMeanData), max(rowMeanData), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))
lines(xfit, yfit, pch=22, col="red", lty=2)

# Add legend to the chart
legend('topright', c("Simulated", "Theoretical"), lty=c(1,2), col=c("black",
"red"))
```



The analytics mean is 4.991311 the theoretical mean 5. The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution

How variable it is and compare it to the theoretical variance of the distribution..

1.standard deviation of distribution

```r
simulatedSd <- sd(rowMeanData) # Standard Deviation
theorySd <- (1/lambda)/sqrt(n) # Standard Deviation
rbind(simulatedSd, theorySd)

##                  [,1]
## simulatedSd 0.8022153
## theorySd    0.7905694
```

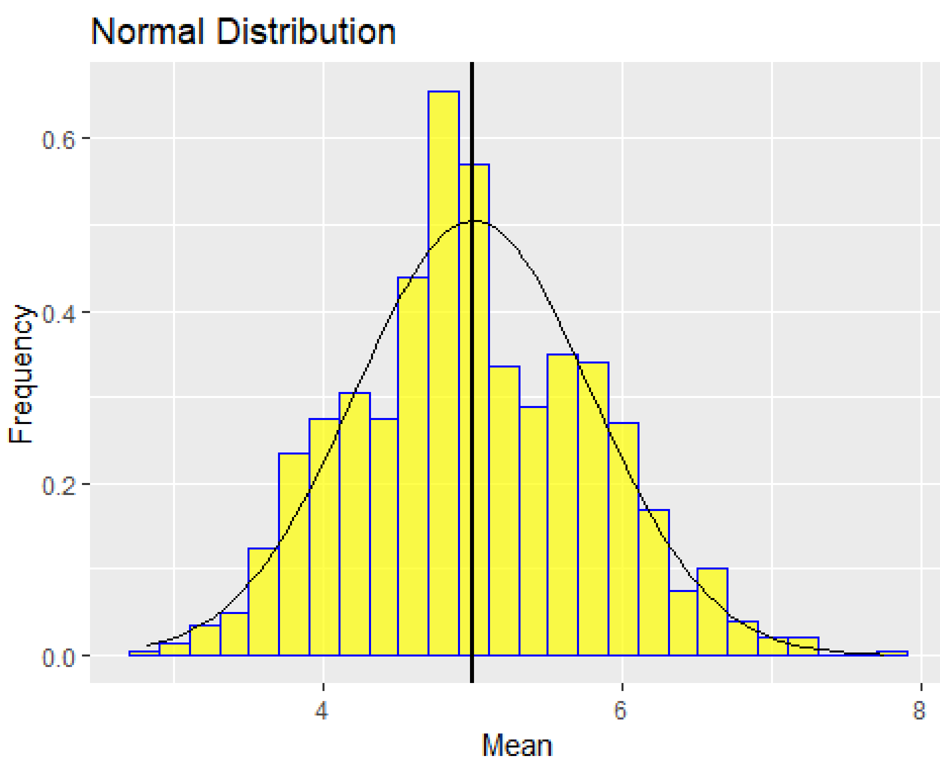
```r
simulatedVar <- simulatedSd^2 # Variance
theoryVar <- ((1/lambda)*(1/sqrt(n)))^2 # Variance
rbind(simulatedVar, theoryVar)
```

```
##                   [,1]
## simulatedVar 0.6435493
## theoryVar    0.6250000
```

Standard Deviation of the distribution is 0.8022153 with the theoretical SD calculated as 0.7905694. The Theoretical variance is calculated as 0.6250000. The actual variance of the distribution is 0.6435493

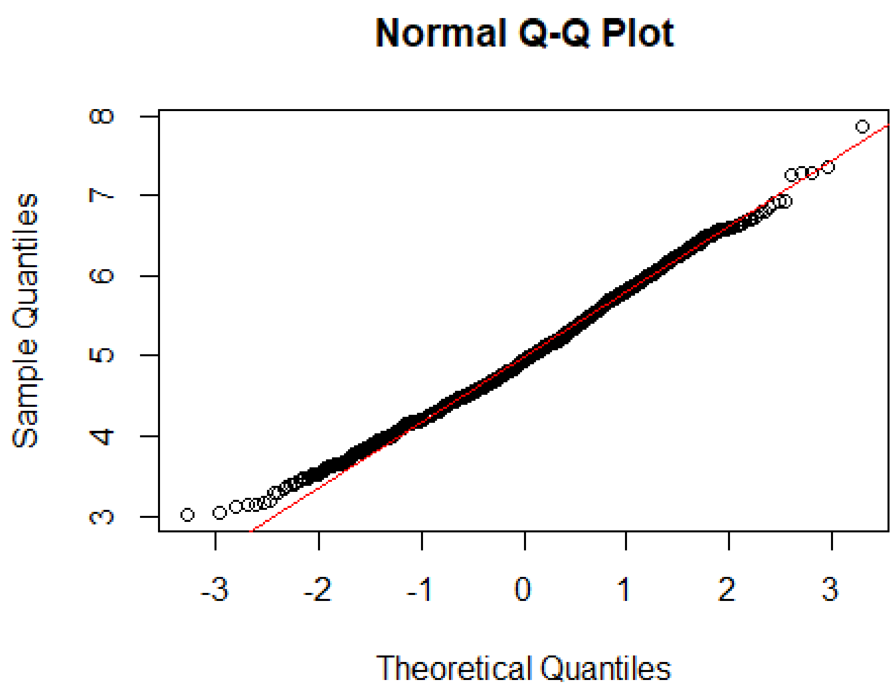
Is the distribution is approximately normal?

```r
data2 <- data.frame(rowMeanData)
hist <- ggplot(data2, aes(x = rowMeanData))
hist <- hist + geom_histogram(aes(y = ..density..), colour = "blue",
                              fill = "yellow", alpha = .7,binwidth=.2)
hist <- hist + stat_function(fun = "dnorm", args = list(mean = theoryMean, sd
= theorySd))
hist <- hist + geom_vline(xintercept=theoryMean,size=1)
hist <- hist + xlab("Mean")+ylab("Frequency")+ ggtitle("Normal Distribution")
hist
```



compare the distribution of averages of 40 exponentials to a normal distribution

```
qqnorm(meanData)
qqline(meanData, col = 2)
```



Since the points fall very close to the line due to Normal Distribution, we can say with some confidence that sample means follow normal distribution